

第12回 重回帰分析と平均値の比較

1. 重回帰分析

第10回の実習課題の発展として、重回帰分析を試みる。第10回の課題が終わっていない人は、第10回の課題をまず済ませてから、以下の課題に取り組むこと。

【1】所得income2を従属変数として、年齢age10と教育年数edctnを独立変数とする重回帰方程式を求め、その意味を解釈しなさい（第10回授業レジュメ参照）。

つぎのような回帰式を考える。

$$\text{INCOME2} = a + b_1 \text{AGE10} + b_2 \text{EDCTN}$$

spss のコマンドは、`regression /dependent =income2 /method enter age10 edctn.`

チェックポイント

(1) a は切片（定数）、 b_1 は EDCTN が一定の場合 AGE10 が 1 単位（10 歳）増加するにつれて、INCOME が何単位（万円）増加するかを意味する。 b_2 は AGE10 が一定の場合 EDCTN が 1 単位（年）増加するにつれて INCOME が何単位（万円）増加するかを意味する。

(2) 決定係数は、この回帰式によって、従属変数の分散の何パーセントが説明されるかを示す。また、決定係数の平方根である重相関係数は、予測された従属変数（所得）と観測された従属変数（所得）の相関係数である。F検定結果の有意水準は、重回帰式による説明力 (R^2) が有意であるかどうかの危険率を示す。

(3) 係数の有意確率は、 a 、 b_1 (β_{a1})、 b_2 (β_{a2}) がそれぞれ有意であるかどうかの危険率を示す。

【2】所得income2を従属変数として、年齢age10と学歴q43を独立変数とする重回帰方程式を求め、その意味を解釈しなさい。

spss のコマンドは、`regression /dependent =income2 /method enter age10 q43.`

※学歴 q43 の 1 単位は、学歴 1 段階（中卒→高卒→短大・高専卒→大卒）であることに注意。

【3】所得income2を従属変数として、年齢age10、教育年数edctn、性別gender(男性= 0、女性= 1)を独立変数とする重回帰方程式を求め、その意味を解釈しなさい。決定係数、重相関係数、標準化された偏回帰係数、分散分析の F 検定結果（有意水準）にも注意すること。

第10回と同じ要領で【1】【2】【3】の結果とその解釈をレポートにまとめ、コースの「重回帰分析」のレポート・ボックスに提出すること。

2. 平均値の比較と相関比

独立変数 x のカテゴリーごとに従属変数 y の平均値を計算する spss のコマンドは、
means 従属変数y by 独立変数x /statistics anova.

言うまでもないが、従属変数は量的変数、独立変数は質的変数でなければならない。
statistics コマンドの anova は分散分析と F 検定結果を出力する。

【例題】回答者の年齢別の平均所得を求め、年齢による所得の格差が有意であるかどうかを検定しなさい。

means income2 by age10/statistics anova.

を実行する。

報告書

各年齢階級ごとの平均所得が示される。

INCOME2				
AGE10	平均値	度数	標準偏差	
1.00 20-29	250.0000	101	202.23748	
2.00 30-39	389.6970	165	303.30984	
3.00 40-49	463.2258	155	364.41010	
4.00 50-59	560.1563	192	395.74104	
5.00 60-69	387.7976	168	366.91369	
合計	427.7209	781	356.77031	

分散分析表から、年齢による所得の格差が統計的に有意である（たんなるサンプリング誤差ではない）ことが分かる

分散分析表

		平方和	自由度	平均平方	F値	有意確率
INCOME2 x AGE10	グループ間 (結合)	7259290.773	4	1814822.693	15.304	.000
	グループ内	92023052.377	776	118586.408		
	合計	99282343.150	780			

年齢と所得との関連の強さは、相関比イータで示される。イータの2乗は決定係数であり、

連関の測定方法

所得の分散の何パーセントが年齢によって説明されるかを示す。この例では、7.3%。

	イータ	イータの2乗
INCOME2 x AGE10	.270	.073

【回答例】年齢階級別の有職者の平均所得は、表1の通り（報告書の表を貼り付け、「表1 年齢階級別の有職者の平均所得」という表題をつける）。相関比は.270 ($p < .001$)。よって、年齢による所得格差は統計的に有意である。

【1】 例題を実行し、その結果をワード文書にまとめること。

【2】回答者の学歴（q43）別の平均所得を求め、学歴による所得格差が統計的に有意であるかどうかを検定しなさい。

【3】回答者の性別（q42a）平均所得を求め、性別による所得格差が統計的に有意であるかどうかを検定しなさい。

以上の3つの課題については、ひとつの文書にまとめて、コーラスの「平均値の比較」のレポート・ボックスに提出すること。